NAVAL HEALTH RESEARCH CENTER

RUNNING ECONOMY:
COMPARING ALTERNATIVE MEASUREMENT MODELS

R. R. Vickers, Jr.

Report No. 05-14

Approved for public release; distribution unlimited.

NAVAL HEALTH RESEARCH CENTER
P. O. BOX 85122
SAN DIEGO, CA 92186-5122

BUREAU OF MEDICINE AND SURGERY (M2)
2300 E ST. NW
WASHINGTON, DC 20372-5300

Running Economy:
Comparing Alternative Measurement Models

Ross R. Vickers, Jr., Ph.D.

Warfighter Performance Department
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122

Report 05-14, supported by the United States Office of Naval Research under Work Unit 63706N M0096.001-60517. The views expressed in this article are those of the author and does not reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government. Approved for public release; distribution unlimited. This research has been conducted in compliance with all applicable federal regulations governing the protection of human subjects in research.

SUMMARY

Background

Running economy (RE) is an efficiency index. Greater economy means that a person uses less energy to run a given distance. Given two equally fit runners, the more economical will run a given distance faster than the less efficient runner. For two runners with different levels of fitness, a less fit runner may be able to use greater economy to outperform a fitter, less economical runner. Accurate measurement of RE is necessary to quantify its effects precisely.

Objective

Standard RE measures are based on the assumption that oxygen consumption rate is a linear function of running speed. The slope of the line (b_1), which is the amount of oxygen needed to run 1 unit of distance, defines RE. If the intercept of the function (b_0) is zero, the oxygen uptake rate measured at any running speed can be divided by speed to obtain an unbiased estimate of b_1. Nevill, Cooke, Holder, Ramsbottom, and Williams (1992) presented evidence that $b_0 \neq 0$. Instead, b_0 varies from person to person. The objective of this study was to further evaluate the bias in RE measures implied by this finding.

Methods

Data came from 5 published studies that reported steady-state oxygen uptake for individual participants who had completed running bouts at 3 or 4 different submaximal speeds. The measured oxygen uptake (mlO_2) was modeled as a linear function of running speed (s) (i.e., $mlO_2 = b_0 + b_1*s$). Analysis of covariance (ANCOVA) and hierarchical linear modeling (HLM) evaluated both the intercepts (b_0) and slopes (b_1) for this model.

Results

Two mathematical models provided reasonable summaries of the data. One model, which combined individual differences in b_0 with a fixed b_1, replicated the Nevill et al. (1992) findings. The other model, which combined a fixed b_0 with individual differences in b_1, indicated that a simple modification would eliminate bias from standard RE measurement methods. The modification would subtract 3.18 $ml0_2 \cdot kg^{-1} \cdot min^{-1}$ from measured oxygen uptake. The second model was only slightly better in explaining the variation in oxygen consumption, but it was supported by HLM findings indicating that b_1 varied significantly across individuals while b_0 did not. Even these modest differences were informative in light of the very strong correlation (r = .935) between individual differences in b_0 and individual differences in b_1.

Conclusions

Logical considerations also favor the second model. For example, the model with variable b_0 would require an explanation for the fact that some individuals had negative resting oxygen uptake rates, a physiological impossibility. Combining logic with the empirical evidence, the conclusion is that a slight modification the standard approach yields unbiased RE estimates. The

RE modification combines individual differences in b_1 with a fixed b_0 value. The combination yields an adjusted RE based on net oxygen uptake. The adjustment based on the present evidence would be 3.18 $ml0_2 \cdot kg^{-1} \cdot min^{-1}$, so adjusted RE = $(mlO_2 – 3.18)/s$. Other research suggests that the proper adjustment may be in the range from 3.6 to 5.0 $ml0_2 \cdot kg^{-1} \cdot min^{-1}$. Further study to better define the required adjustment could be useful. A review of existing research found adjusted RE measures in about 1 of every 12 studies. The present findings suggest that this practice should be the standard approach instead of a rarity.

Introduction

The oxygen cost of running (OCR) is the oxygen consumed in running a given distance. This cost, which is typically expressed as $mlO_2 \cdot kg^{-1} \cdot km^{-1}$ or $mlO_2 \cdot kg^{-1} \cdot m^{-1}$, is assumed to be a fixed value determined by the distance covered. OCR is assumed to be independent of the speed at which the distance is covered. These assumptions are based on a linear relationship between the rate of oxygen uptake and running speed for speeds above 8 $km \cdot hr^{-1}$ (McArdle, Katch, & Katch, 2001).

Individual differences in OCR are a common research topic. Running economy (RE) is the label given to these differences. An individual with a lower value for OCR (i.e., greater economy) will expend less cumulative energy to run a given distance than will an individual with higher OCR. The RE label highlights this difference in efficiency and directs attention to an important attribute. RE can explain why performance is imperfectly linked to maximal oxygen uptake capacity. Given two runners with equal maximal capacity, the more economical of the two will be able to convert this capacity into faster run times. Given two runners with different maximal capacities, the one with less capacity may perform better than his or her counterpart if he or she is more economical.

The belief that OCR is independent of running speed simplifies the measurement of RE. A person can run at a submaximal speed long enough to reach steady-state oxygen uptake. The RE estimate is the steady-state uptake divided by the running speed. If OCR is independent of speed, this ratio will be the same no matter what running speed is chosen for the test. The only requirement is that the speed should be >8 $km \cdot hr^{-1}$, the approximate speed at which the preferred mode of locomotion changes from walking to running (McArdle et al., 2001).

The belief that OCR is independent of running speed is not likely to be literally correct. The general equation for a linear relationship is $y = b_0 + b_1x$. Substituting mlO_2 as the dependent variable and speed (s) as the predictor, this general equation becomes $mlO_2 = b_0 + b_1s$. OCR is independent of speed only if $b_0 = 0$. In this special case, $mlO_2 = b_1 * s$. Because $OCR = mlO_2/s$, the special case of the general equation can be rewritten to yield $OCR = mlO_2/s = b_1$. The last term in this sequence of equalities does not include speed, so OCR has the same value (i.e., b_1) regardless of speed. Thus, OCR is independent of speed only in the special case of $b_0 = 0$. Empirical instances of the special case would be an exception to the more common finding that $b_0 \neq 0$ (c.f., Leger & Mercier, 1984; Nevill et al., 1992).

The bias in standard RE measures when $b_0 \neq 0$ is readily evident. In this case, $mlO_2/s = b_1 + b_0/s$. The second term on the right hand side of the equation, b_0/s, is the bias. Note that bias does not necessarily affect the assessment of RE. If $b_0 = k$ for all individuals, bias is the same for all individuals. In this case, bias will affect the average cost, i.e., OCR, but the variation of individual values about that average will not be affected. Thus, standard OCR measurements translate into proper RE estimates if $b_0 \approx k$ for all individuals.

Nevill et al. (1992) presented evidence that b_0 differs substantially between individuals. In their study, each participant performed 3 submaximal runs at different speeds. Steady-state mlO_2 was measured at each speed. The data were analyzed by ANCOVA. Oxygen uptake rate was the dependent variable. Participant was the group variable. Speed was the covariate. The pooled regression of mlO_2 on speed accounted for 79% of the variance. Individual differences in b_0 accounted for 18% of the VO_2 variance.

Nevill et al.'s (1992) findings raise doubts about the status of the RE as a theoretical construct. In their analysis, RE would be evident as individual differences in b_1. These differences accounted for only 0.4% of the variance in mlO_2. Although statistically significant (*p* < .023), considerations of parsimony would exclude RE from the resulting model. Parsimony can be defined as the number of parameters in a model (Popper, 1959). In this case, including the individual differences in b_1 that define RE would approximately double the number of parameters to increase explanatory power by less than 1%. This trade-off in complexity versus explanatory power is not sufficient to justify including RE in the final model.[1]

The model implicit in standard RE measurement practices differs markedly from the model implied by Nevill et al.'s (1992) findings. The difference can be described by considering the pattern of lines that would be expected if the mlO_2 were plotted as a function of speed for each individual in a sample. The standard method assumes that individual differences would be expressed as differences in the b_1 values for the lines. If all of the lines converged to zero when $s = 0$, these differences in b_1 would yield a fan-shaped array of lines. This possible outcome is referred to as the fan model in the remainder of this paper.

Nevill et al.'s (1992) findings imply a very different picture. The variance explained by differences in b_1 is so slight that slope can be regarded as a constant for all participants. The variance explained by differences in b_0 is relatively large, so a plot of the data for each individual would show a set of parallel lines. The lines would be parallel because they would all share the common value for b_1, but differences in b_0 would separate the lines for different subjects. The differences in b_0 might have a physiological interpretation. These differences might, for example, be equated with differences in resting metabolic rate or resting energy expenditure. These suggestions are only illustrative, because b_0 differences really refer to estimated mlO_2 associated with a hypothetical behavior (i.e., running at $s = 0$). This behavior, if possible at all, might not equate with any current physiological construct. Still, Nevill et al.'s findings imply that basal uptake differences are the primary source of individual differences in oxygen uptake while running. This perspective is referred to as the parallel lines model in the remainder of this paper.

Nevill et al.'s (1992) findings supported the parallel lines model over the fan model. This study focused on 2 factors that might alter the choice of models. First, Nevill et al.'s (1992) findings, like those of any study, could be the product of chance. Second, their statistical procedures may have been biased in favor of the parallel lines model. ANCOVA procedures often involve a sequential decomposition of variance. The sequence determines, in order, the

[1]Individual differences in b_1 did produce a statistically significant increase in the explanatory power of the model. However, statistical significance is a poor guide for modeling decisions in this case. Statistical significance meant that the variance explained by differences in b_1 was greater than expected by chance. The outcome was obtained only because the overall model accounted for almost all of the mlO_2 variance. The small residual variance translated into a small denominator for the F-test that was the basis for determining statistical significance. Because the error term was small relative to the overall variance, even trivial explanatory power could yield a large enough F-test to classify the explanatory power as greater than expected by chance. The fact that differences in b_0 accounted for 45 times as much variance as differences in b_1 is a better barometer of the importance of b_1 variation.

variance explained by a common regression line, by group differences in the intercept of the regression line, and by group differences in the slope of the regression line.[2] This sequence will affect inferences in the present case if differences in b_1 correlate with differences in b_0. Correlated differences imply overlapping explanatory power. The default sequence assigns any overlapping explanatory power to the b_0 differences. From a modeling perspective, there is no a priori reason why the variance should be partitioned in this manner. Some or all of the overlapping explanatory power could be attributed to b_1. Any reallocation would obviously affect conclusions about the importance of b_1 differences. The implication is that Nevill et al.'s (1992) statistical model may have been biased implicitly against retaining the RE construct.

Table 1. Descriptions of Samples

	N	Gender	Age	Weight (kg)	VO_2max	# of Speeds	Range of Speeds
A	16	♂	35	63.1	66.6	3-4	215-322
B_1	8[a]	♂	18.4	65.8	59.7	3-4	249-360
B_2	6	♀	16.5	51.5	47.0	4	220-290
C_1	4	♂	32	68.8	69.3	3	214-268
C_2	4	♀	25	54.0	59.2	3	188-241
D_1	6	♂	10	30.4	NA[b]	3	133-183
D_2	6	♂	33	77.4	NA[b]	3	133-183
E	10	♂	26	NA[b]	61.7	3	161-210

Note. Studies were A = Costill, Thomason, & Roberts, 1973; B = di Prampero et al., 1993; C = Morgan et al., 1994; D = Thorstensson, 1986; E = Williams, Krahenbuhl, & Morgan, 1991. Blank spaces indicate that the information was not reported.
[a]Two males who ran at only two velocities were dropped.
[b]Information not available.

This paper examines both possible explanations for the Nevill et al. (1992) findings. The examination, which relies on published data from 5 studies, establishes 3 primary points:

- Nevill et al.'s findings replicate.
- Individual differences in b_1 are strongly correlated with individual differences in b_0.
- Oxygen uptake can be predicted about equally well by a model with differences in either b_0 or b_1.

The discussion examines the reasons for preferring the fan model to the parallel lines model in light of these findings. Because the fan converges at a non-zero value for the common b0, the discussion also considers the modification of standard RE measurement methods.

[2]Nevill et al. (1992) did not explicitly state that their analysis employed the sequential sums of squares (SS) approach. However, the total SS for the model equals the sum of the SS for the pooled regression, individual differences in b_0, and individual differences in b_1. This equivalence is unlikely if alternatives to the sequential SS procedure (e.g., unique SS) were used.

Methods

Data Sources

The PubMed database was searched with "running economy" and with "oxygen cost AND running" as search terms. The reference lists from articles identified in the PubMed search were reviewed to identify additional candidate studies. Table 1 describes the 5 studies (Costill et al., 1973; di Prampero et al., 1993; Morgan et al., 1994; Thorstensson, 1986; Williams et al., 1991) that reported the data for individuals who completed submaximal runs at 3 or more speeds. The 60 participants in those studies formed the sample for the present analyses.

Experimental Procedures

The methods used in the studies were broadly similar. The studies all had participants perform 3 or 4 runs at submaximal speeds. Each run lasted long enough to ensure that oxygen uptake reached steady-state. The oxygen consumption rate was measured for 2 min at the end of the run in 4 studies. The measurement period varied from 28 sec to 70 sec in the fifth study (di Prampero et al., 1993). Table 1 provides additional descriptive data for the studies.

Analysis Procedures

The analyses had 4 major components. The first component was replication. Nevill et al.'s (1992) ANCOVA procedures were repeated separately for each of the 8 samples described in Table 1. The general linear model (GLM) procedure in SPSS-PC (SPSS, Inc., 1998a, 1998b) provided this element of the analysis. Individual subjects were the "group" variable; running speed was the covariate. The ANCOVA model employed the sequential decomposition of variance that appeared to have been used in the Nevill et al. analyses.

The second analysis component introduced an alternative to the Nevill et al. (1992) model. The alternative model was defined by changing the order of entry of predictors in the ANCOVA. The model specification option of the GLM procedure was used to add individual differences in b_1 first followed by individual differences in b_0.

Statistical significance and explanatory power were both examined as criteria for evaluating the results of these first 2 components of the analyses. Explanatory power was expressed as variance explained adjusted for the number of parameters fitted. The adjustment produces the ω^2 statistic (Hays, 1963) instead of ε^2. In an ANOVA, ε^2 is the variance explained by a factor divided by total variance (i.e., SS_{Factor}/SS_{Total}). This index of explanatory power includes some variance that can be attributed to chance. Even if the factor being examined had no effect on the dependent variable, the expected amount of variance explained would be greater than zero. In particular, the apparent explanatory power would equal the degrees of freedom for the factor multiplied by the mean squared error (i.e., $df*MS_{res}$). The ω^2 statistic estimates explanatory power adjusting for chance differences. In particular, $\omega^2 = [SS_{Factor}-(df*MS_{res})]/SS_{Total}$. If $MS_{res} > 0$, then $\omega^2 < \varepsilon^2$. However, ω^2 arguably is an estimate of the systematic variance explained by the factor. The distinction between variance attributable to chance and systematic variance was important in the present analyses because of the large number of degrees of freedom associated with individual differences in the slope and the intercept. The raw variance explained by those differences could be substantial even if these factors essentially explained chance amounts of variation.

The third analysis component examined the degree of confounding of individual differences in b_0 and b_1. Linear regression equations were computed for each individual in each sample. The b_0 and b_1 parameter estimates were recorded for each person. The Pearson product-moment correlation, *r*, was computed to quantify the covariation of these parameters.

The fourth analysis component was a set of hierarchical linear models performed with the HLM5 computer program (Raudenbush, Bryk, Cheong, & Congdon, 2001). The predictors in these models were running speeds nested within subjects, subjects nested within studies, and studies. The multilevel models tested the hypotheses that b_0 and b_1 varied significantly between individuals after controlling for study differences (cf., Raudenbush & Bryk, 2002). The model was computed with speed centered about the mean running speed which was 233.08 $m \cdot min^{-1}$. Centering is routinely used in multilevel modeling to reduce confounding between predictors. In this case, it was necessary to reduce the confounding of b_0 and b_1 estimates. The effects of centering must be taken into account when interpreting the data.

Results

Nevill et al.'s (1992) findings generally replicated when their analysis methods were followed. The relevant results are shown in columns A, B, and D_1 of Table 2. Column A indicates that the pooled regression line accounted for ~83% of the variance in mlO_2 (range = 54% - 92%). Column B indicates that individual differences in b_0 accounted for an additional ~11% of the variance (range = 3% to 28%). Individual differences in b_1 added an average of ~1% of the variance (range = 0% - 2%) to the explanatory power of the model (Column D_1).

One aspect of the Nevill et al. (1992) findings did not replicate. The variance explained by individual differences in b_1 was statistically significant ($p < .05$) in only 1 of 8 samples (Williams et al., 1991). The variance explained in this case actually was less than average (0.4%). The overall model accounted for nearly all of the variance, so the small residual made even weak effects significant. The same was true for Nevill et al. (1992), so b_1 differences had little explanatory value even when they were statistically significant.

Table 2. Summary of ANCOVA Results

	Ω^2 for Model:					Significance:	
	A	B[a]	C[b]	D_1[c]	D_2[d]	D_1	D_2
Costill	.916	.033	.034	.013	.012	.069	.086
di Prampero	.779	.053	.060	.020	.013	.223	.289
di Prampero	.546	.283	.297	.024	.010	.200	.323
Morgan	.808	.176	.179	.001	.000	.414	.754
Morgan	.784	.206	.208	.004	.001	.146	.308
Thorstensson	.870	.113	.111	.000	.000	.630	.454
Thorstensson	.873	.066	.061	.007	.011	.340	.271
Williams	.863	.131	.130	.004	.005	.000	.000
Nevill	.792	.182	.	.004	.	.023	.
Average	.827	.109	.112	.010	.008		

Note. Model A = Common regression line; Model B = Individual differences in intercepts; Model C = Individual differences in slopes; Model D = Individual differences in slopes *and* intercepts.
[a]Relative to null model.
[b]Relative to Model A.
[c]Relative to Model B.
[d]The actual value (–0.15) presumably represented the effects of chance sampling variation about a true value of 0.00.
[e]Nevill et al. (1992) excluded to base comparisons on the same set of observations.

Table 2 also describes the results of an alternative sequence suitable for testing the fan model. This sequence began with the same pooled group regression (Column A). Adding differences in b_1 at the second step accounted for ~11% of the variance (Column C). Adding differences in b_0 as the third model element accounted for ~0.8% of the variance. The b_0 differences had statistically significant explanatory power in only 1 of the 8 samples (Williams et al., 1991).

Confounding of Individual Differences

The results summarized in Table 2 were consistent with the suggestion that individual differences in b_1 correlate with individual differences in b_0. The combined explanatory power of b_0 and b_1 differences was ~12%. Almost all of this explanatory power (11%) could be accounted for by either model component. This clear and substantial overlap in explanatory power can be explained by the strong association between b_0 and b_1 (r = -.935). The association was slightly stronger in the HLM analysis with uncentered data (r = -.961). The difference between these two estimates of the degree of overlap probably can be attributed to the fact that the HLM analyses included an adjustment for the effect of differences between studies.

HLM Results

A fixed-effects model is one element of hierarchical linear models. In the present case, this is the model that would be adopted if there were no evidence of individual or study differences in the regression parameters. This element of the model was the equation:

$$mlO_2' = 45.30 + .208*(s - 233.08)$$

The subtraction of 233.08 m/min (~14 $km \cdot hr^{-1}$) from speed centered the analysis on the grand mean. The intercept in the centered analysis is the predicted mlO_2 at the mean speed. The intercept for the raw data can be obtained by solving the equation for s = 0. The raw data intercept was 3.18 $ml \cdot kg^{-1} \cdot min^{-1}$.

The random-effects components of the model indicated that individual and study were significant sources of variation in the regression coefficients. Individual differences in b_1 and b_0 were random-effects components of the model. Differences in b_1 were significant ($\chi^2 = 81.61$, 55 *df*, $p < .012$). Differences in b_0 only approached significance ($\chi^2 = 69.27$, 55 *df*, $p < .094$). Study differences were significant for both b_0 ($\chi^2 = 81.33$, 4 *df*, $p < .001$) and b_1 ($\chi^2 = 20.68$, 4 *df*, $p < .002$).

Discussion

The present analyses are consistent with a modified fan model for individual differences in mlO_2 at submaximal running speeds. This model is preferable to the parallel lines model implied by Nevill et al.'s (1992) findings. The fan model would be modified to include a non-zero value for mlO_2 when s = 0. The following discussion elaborates on these basic points.

The conclusion that a modified fan model is appropriate rests on several considerations. The fact that Nevill et al.'s (1992) findings only partially replicated is one factor. With regard to that replication, the present analyses showed that a common regression model that applied to all participants was a good first approximation model in every study. The replication extended further to the demonstration that adding individual differences in b_0 produced a substantial average improvement in model accuracy. The replication was completed by evidence that adding individual differences in b_1 produced only a trivial improvement in predictive accuracy. While this last increment of variance explained was statistically significant in Nevill et al.'s study, it was statistically nonsignificant in 7 of 8 samples analyzed here. The consistent weakness of the b_1 effects argued for a model with a common regression slope, but different intercepts. In other words, the evidence supported a parallel lines model when the analyses replicated the Nevill et al.'s procedures.

The examination of the alternative statistical model gave reason to consider rejecting the parallel lines model. Analyses guided by the fan alternative showed that substituting differences in b_1 for differences in b_0 produced slightly better predictive accuracy (~0.3%) at the second step in the model. Adding differences in b_0 at the third step produced a trivial gain in predictive accuracy. The predictive power of the fan and parallel lines models was nearly identical because differences in b_0 and b_1 were very strongly related ($r = -.935$).

The fan and parallel lines models were the only viable alternatives. These models were roughly equivalent when evaluated by the combined criteria of explanatory power and simplicity. Other alternative models were less satisfactory by these criteria. A model in which both b_0 and b_1 varied would require many more parameters with a trivial gain in explanatory power. A model with both b_0 and b_1 assigned fixed values that applied to all individuals would be less complex (i.e., only 2 parameters), but it would have about 11% less explanatory power than either the fan or parallel lines models.

The fan model is preferable to the parallel lines model. The HLM analyses provided the strongest empirical basis for preferring the fan model. The variance of b_1 across subjects was large enough to reject the null hypothesis at the traditional $p < .05$ significance level. The variance in b_0 approached but did not reach this significance criterion. The slightly greater average predictive accuracy (~0.3%) is a secondary empirical justification for this preference.

The fan model also is preferable to the parallel lines model on conceptual grounds. Differences in b_1 correspond to the conceptual definition of RE. Differences in b_0 lack a clear correspondence to any existing construct. These differences might be equated with basal metabolism or resting energy expenditure, but there is good reason to dismiss these possibilities. The analyses indicated that 60% of the b_0 estimates were negative; 40% were $\leq$ -5 $ml \cdot kg^{-1} \cdot min^{-1}$. These negative b_0 estimates do not have a plausible physiological interpretation. Other things equal, a model that includes a parameter, b_1, that corresponds to a physiologically meaningful construct, RE, is preferable to one that requires some convoluted explanation for cases in which $b_0 < 0$.

Adopting the fan model retains the RE construct as a legitimate concept, but the current form of the model also implies that standard methods of measuring RE should change. RE measures should adjust for b_0 by subtracting 3.18 $ml \cdot kg^{-1} \cdot min^{-1}$ from the measured mlO_2. In a review of studies that measured RE in 201 samples, 8% of the studies used a fixed-value adjustment, usually 5 $ml \cdot kg^{-1} \cdot min^{-1}$ based on Medbo, Mohn, Tabata, Bahr, Vaage, and Sejersted (1988). The present findings therefore support a shift in emphasis rather than an entirely novel approach to RE.

RE measures based on net oxygen uptake (i.e., the adjusted fan) raise a number of research questions.

- Are individual differences in b_0 really the product of chance? Repeated measures designs such as those used by Morgan et al. (1994) and Williams et al. (1991) could answer this question. The data from these designs could be used to compute b_0 and b_1 on multiple occasions. Stable individual differences in b_0 would argue that these differences are not the results of chance.
- If b_0 is a constant, what is the correct value? Values of 3.18 $ml \cdot kg^{-1} \cdot min^{-1}$ and 5.0 $ml \cdot kg^{-1} \cdot min^{-1}$ were mentioned previously; published regression equations relating mlO_2 to running speed (Leger & Mercier, 1984; Nevill et al., 1992) provide a range of other options. Research is needed to better define the required adjustment.
- Does adjustment increase validity? Studies that compared adjusted and unadjusted RE measures as correlates of running performance or other relevant variables could answer this question. No such studies were identified in a recent extensive review of the relevant literature.

Study design effects on RE also merit further study. The HLM analyses showed large differences between studies. Each study is a complex configuration of factors that could contribute to these differences. A sample of 5 studies is too small to explore this variation. Further study would be worthwhile to determine what factors affect the RE estimates.

This replication and extension of Nevill et al.'s (1992) inquiry leads to the conclusion that RE should be estimated from the net oxygen cost of running. The net cost approach subtracts a basal value for oxygen uptake from the measured value. The net value arguably reflects the actual cost of the activity involved in running (McArdle et al., 2001). The adjustment will yield unbiased RE estimates. Studies comparing the validity of adjusted and unadjusted RE estimates would be a very useful extension of this investigation.

References

Costill, D. L., Thomason, H., & Roberts, E. (1973). Fractional utilization of the aerobic capacity during distance running. *Medicine and Science in Sports, 5*(4), 248-252.

di Prampero, P. E., Capelli, C., Pagliaro, P., Antonutto, G., Girardis, M., Zamparo, P., et al. (1993). Energetics of best performances in middle-distance running. *Journal of Applied Physiology, 74*(5), 2318-2324.

Hays, W. L. (1963). *Statistics for psychologists*. New York: Holt, Rinehart, Winston.

Leger, L., & Mercier, D. (1984). Gross energy cost of horizontal treadmill and track running. *Sports Medicine, 1*(4), 270-277.

McArdle, W. D., Katch, F. I., & Katch, V. L. (2001). *Exercise physiology: Energy, nutrition, and human performance* (Fifth ed.). Philadelphia: Lippincott Williams and Wilkins.

Medbo, J. I., Mohn, A. C., Tabata, I., Bahr, R., Vaage, O., & Sejersted, O. M. (1988). Anaerobic capacity determined by maximal accumulated O_2 deficit. *Journal of Applied Physiology, 64*(1), 50-60.

Morgan, D. W., Craib, M. W., Krahenbuhl, G. S., Woodall, K., Jordan, S., Filarski, K., et al. (1994). Daily variability in running economy among well-trained male and female distance runners. *Research Quarterly for Exercise and Sport, 65*(1), 72-77.

Nevill, A. M., Cooke, C. B., Holder, R. L., Ramsbottom, R., & Williams, C. (1992). Modeling bivariate relationships when repeated measurements are recorded on more than one subject. *European Journal of Applied Physiology, 64*, 419-425.

Popper, K. R. (1959). *The logic of scientific discovery*. New York: Basic Books.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S., Bryk, A., Cheong, Y. F., & Congdon, R. (2001) *HLM 5: Hierarchical linear and nonlinear modeling*. Chicago: Scientific Software International.

SPSS, Inc. (1998a). *SPSS Advanced Statistics*. Chicago: SPSS, Inc.

SPSS, Inc. (1998b). *SPSS Base 8.0 Applications Guide*. Chicago: SPSS, Inc.

Thorstensson, A. (1986). Effects of moderate external loading on the aerobic demand of submaximal running in men and 10 year-old boys. *European Journal of Applied Physiology, 55*, 569-574.

Williams, T. J., Krahenbuhl, G. S., & Morgan, D. W. (1991). Daily variation in running economy of moderately trained male runners. *Medicine & Science in Sports & Exercise, 23*(8), 944-948.

REPORT DOCUMENTATION PAGE

1. Report Date (DD *MM YY*): 2 Mar 2005

2. Report Type: final

3. DATES COVERED (from - to): 10/01/04 – 03/01/05

4. TITLE AND SUBTITLE
Running Economy: Comparing Alternative Measurement Models

5a. Contract Number:
5b. Grant Number:
5c. Program Element: 63706N
5d. Project Number: M0096
5e. Task Number: 001
5f. Work Unit Number: 60517
5g. IRB Protocol Number: NA

6. AUTHORS
Vickers, Ross R., Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122

8. PERFORMING ORGANIZATION REPORT NUMBER
Report No. 05-14

8. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

Office of Naval Research
800 N. Quincy St.,
Arlington, VA 22217-5660

Chief, Bureau of Medicine and Surgery
Code M53
2300 E Street N.W
Washington DC 20372-5300

10. Sponsor/Monitor's Acronyms(s)
BuMed/ONR

11. Sponsor/Monitor's Report Number(s)

12 DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT (maximum 200 words)

The oxygen cost of running (OCR) is the oxygen consumed per unit of distance. Running economy (RE) refers to individual differences in OCR. OCR and RE are measured by dividing the oxygen uptake (mlO_2) rate during a submaximal run by the speed (s) of that run. The results of these measurements are assumed to be independent of speed. This assumption is invalid unless the intercept (b_0) of the regression of mlO_2 on s is zero. When $b_0 \neq 0$, standard measures are biased. The magnitude of bias is b_0/s, so bias is contingent on speed. Nevill, Cooke, Holder, Ramsbottom, and Williams (1992) presented evidence that, in general, $b_0 \neq 0$. Instead, b_0 differs between people, so bias is a problem that may obscure important relationships in RE research. This report presents a re-analysis of published data from 5 studies to replicate Nevill et al.'s (1992) findings and examine plausible alternatives to their model. Nevill et al.'s findings did replicate, but an alternative mathematical model that replaced differences in b_0 with differences in b_1, the slope of the regression of mlO_2 on s, was preferable. The alternative model was slightly more accurate in predicting mlO_2 and the variation in individual differences in b_1 was significant while the variation in differences in b_0 was not. The alternative model also was preferable because differences in b_1 were interpretable in bioenergetic terms. A high proportion of negative values made the interpretation of differences in b_0 uncertain. Because the alternative model included a fixed b_0, it implied that RE measures should be based on net oxygen consumption. Subtracting ~3.2 $ml \cdot kg^{-1} \cdot min^{-1}$ from measured $ml0_2$ would eliminate a speed-related bias. This approach would not be entirely novel. Roughly 8% of RE studies currently use a net uptake approach. The present findings suggest that this practice should become the preferred method of assessing RE

14. SUBJECT TERMS
aerobic capacity, running economy, measurement methods, statistical model, physical fitness

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGE	18a. NAME OF RESPONSIBLE PERSON
a. REPORT	b.ABSTRACT	C. THIS PAGE			Commanding Officer
UNCL	UNCL	UNCL	UNCL	10	18b. TELEPHONE NUMBER (INCLUDING AREA CODE) COMM/DSN: (619) 553-8429

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18